DOCUMENT RESUME

ED 209 285 TM 810 627

AUTHOR Allen, Russell H.; Kaufman, B. Darwin
TITLE Improving Wisconsin Public School Testing Programs.
INSTITUTION Wisconsin State Dept. of Public Instruction, Madison. Div. for Management and Planning Services.
REPORT NO WSDPI-Bull-1376
PUB DATE Mar 81
NOTE 24p.; Tables on pages 3 and 16 marginally legible because of small print.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Program Improvement; *School Districts; Scoring; *Testing Programs; Test Interpretation; Test Selection
IDENTIFIERS *Wisconsin; Wisconsin Pupil Assessment Program

ABSTRACT

The purpose of this paper is twofold: (1) to discuss the variety of factors which should be considered in establishing a testing program; and (2) to summarize some of the basic features of testing programs in Wisconsin public school districts. An effective district-wide testing program should consist of at least the following seven elements: determining a purpose for testing, establishing information requirements, selecting or developing a test, administering the test(s), scoring and analyzing test results, preparing and publishing a report, and providing for follow-up. Wisconsin school districts currently use tests purchased from commercial testing companies, tests purchased from the Wisconsin Pupil Assessment Program, and district-developed tests. The most frequently stated use of test results is to diagnose student strengths and weaknesses. Other frequent uses are to identify curricular strengths and weaknesses, to assist in guidance and counseling, and to place pupils. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED209245

IMPROVING WISCONSIN PUBLIC SCHOOL TESTING PROGRAMS

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

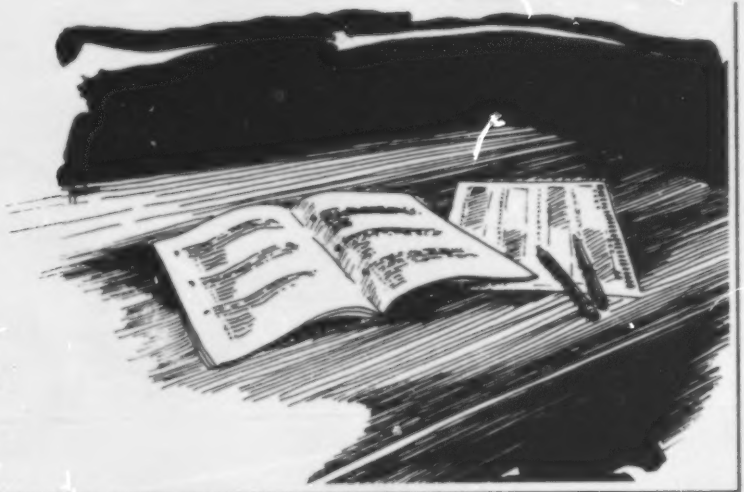
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

L. Harmetuen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."



TM 800 623

Wisconsin Department of Public Instruction
Barbara Thompson, Ph.D., State Superintendent

IMPROVING WISCONSIN PUBLIC SCHOOL TESTING PROGRAMS

by

Russell H. Allen and B. Darwin Kaufman
Wisconsin Pupil Assessment Program

Published by

Division for Management, Planning, and Federal Services
A. A. Buchmiller, Assistant Superintendent
Bureau for Evaluation, Planning, Assessment, and Research
Tom Stefonek, Director

March 1981

Bulletin No. 1376

Introduction

During the past several years, professional educators have faced increased pressure to show they are meeting students' needs for quality education. The public, local boards of education, and elected officials are demanding that schools provide evidence of success. While it would be difficult to list all the factors which have contributed to this demand for accountability in education, the following reasons are frequently cited: long-term declines in student scores on some standardized tests; an increasing dropout rate; the problem of school discipline; increased costs of education; and a strongly and widely held belief that students are not learning the "basics".

The demand for accountability has prompted a few school districts to implement a testing program tied to specific instructional goals and objectives in all subjects and grades. Teaching strategies are designed to reach these goals, and objectives and tests are administered to determine mastery. In contrast, most districts have responded with only the occasional administration of a commercial test.[1]

The purpose of this paper is twofold: (1) to discuss the variety of factors which should be considered in establishing a testing program, and (2) to summarize some of the basic features of testing programs in Wisconsin public school districts

Part I summarizes the variety of factors associated with testing programs. Generally, the discussion of these factors is not detailed; however, there is sufficient information to provide the reader with a basic understanding of what is involved in the implementation and maintenance of a testing program.

Part II presents the results of a questionnaire sent to Wisconsin public school districts in early 1980. The questionnaire was designed to gather basic information about the nature of testing in Wisconsin public schools. The narrative accompanying the various tables presented in Part II is brief because the data need little or no explanation.

---

[1]A testing program represents only one facet of a larger evaluation design. For a relatively complete discussion of educational evaluation, the reader is directed to the *Program Evaluation Kit* by Carol Taylor Fitz-Gibbon and Lynn Lyons Morris, Sage Publications (Beverly Hills, California: 1978).

PART I

ESTABLISHING A TESTING PROGRAM

Establishing an effective testing program requires more than merely purchasing and administering a commercial test. An effective district-wide testing program should consist of at least the following seven elements:

1. Determining a purpose for testing
2. Establishing information requirements
3. Selecting or developing a test
4. Administering the test (s)
5. Scoring and analyzing test results
6. Preparing and publishing a report
7. Providing for follow-up

1. Determining a Purpose for Testing

The initial step in creating a testing program is to define the purpose or purposes of testing (i.e., how the tests will be utilized in making decisions). Unless there is a clear statement of purpose, the existing program will not be focused and may be of little value. Listed below are eight purposes for testing students:

1. identification of curricular strengths and weaknesses
2. promotion or graduation
3. placement of pupils into groups or programs
4. diagnosis of individual student strengths and weaknesses
5. research and development
6. guidance and counseling
7. evaluation of teachers
8. comparison of students to other students

It is important to emphasize that each of the eight purposes is associated with the need to make decisions. In the initial stage of determining the purposes for testing, the decisions to be made should be stated explicitly. For example, the purpose of identifying curricular strengths or weaknesses might be to decide if changes in the curriculum are necessary, and, if so, where they might best be made.

Shown on the next page is a summary table developed by Anderson, Stiggins, and Hiscox, who suggest there are three basic purposes for testing students:

1. to make instructional management decisions,
2. to make decisions about screening students,
3. to make program decisions.[2]

---

[2]Beverly L. Anderson, Richard J. Stiggins, and Suzanne B. Hiscox, *Guidelines for Selecting Basic Skills and Life Skills Tests*, Clearinghouse for Applied Performance Testing (Portland, Oregon: 1980), p. 4.

The Anderson, Stiggins, and Hiscox categorization is of particular value because it describes the relationships among general purposes (assessment context), related decisions, necessary data, and appropriate decision-makers. As a case in point, diagnosis of student strengths and weaknesses is one of the decision areas listed under Instructional Management. Because the purpose of testing in this case is to make decisions about individual students, any test which is adopted must provide information about specific skills and/or subject matter understanding.

Summary of Various Purposes for Testing

Assessment Context	Types of Decisions	Type of Data Needed	DECISION MAKERS: Students	Parents	Teachers	Administrators	Counselors	Public
Instructional Management Decisions								
Diagnosis	Decide students' strengths and weaknesses on specific skills	Individual student data on level of development of specific skills	✓		✓			
Placement	Place student into next most appropriate level of instruction	Scores that place students on relevant knowledge or skill continuum			✓	✓	✓	
Guidance & Counseling	Decide probability of success and satisfaction in given program of educational or vocational development	Data reflecting level of educational development of individual student relative to other students	✓	✓			✓	
Student Screening Decisions								
Selection	Decide which students to be selected into or out of a program	Data that rank order individual students on relevant knowledge or skill scale			✓	✓	✓	
Certification	Determine mastery or non-mastery of specified body of knowledge or set of skills	Data reflecting individual student mastery of specified body of knowledge or set of skills			✓	✓		✓
Programmatic Decisions								
Survey Assessment	Make educational policy decisions; determine educational development of student group	Group achievement data gathered cyclically to show trends				✓		✓
Formative Evaluation	Decide program components in need of modification	Interim and final program outcomes attained and not attained by participating students considered as a group			✓	✓		
Summative Evaluation	Determine if program is to be adopted, expanded, or discontinued	Program outcomes attained and not attained by participating students considered as a group			✓	✓		✓

In the important first step of defining purpose, there are two suggestions. First, there should be a cooperative effort among those who will be required to make decisions based on the test results. This will probably include teachers, administrators, curriculum specialists, counselors, psychologists, and evaluation specialists. Second, it should be recognized that any testing program must be limited in scope. It is unrealistic to propose a testing program which will answer all possible questions. Thus, there is the need to be selective throughout the process of establishing a testing program.

The discussion of purpose offered in this paper is not meant to be exhaustive. Rather, it is intended to alert the reader to the variety of reasons for testing and to the critical need to define the intended purpose or purposes for testing.

2. Establishing Information Requirements

After determining the purpose (s) for testing, the next step is to identify the type of information needed to make the desired decisions. The information provided by any testing program can be classified into one or more of three categories or domains:

1. The affective domain, which refers to attitudes, feelings, interests and values.[3]
2. The psychomotor domain, which refers to those skills involving neuromuscular coordination. Handwriting skills and athletic skills represent two types of psychomotor skills.
3. Cognitive skills, which are identified in a variety of ways. They are usually classified along a continuum. Knowledge (that is, knowledge of facts, rules, and sequences) is viewed as a "lower" order cognitive skill, while "higher" order skills include the ability to classify, to recognize relationships, to analyze, to synthesize, to evaluate, etc.

If, for example, the purpose for testing is to make a decision about the placement of students into academic classes of differing ability levels, it will be necessary to collect data which yield information about how students

---

[3]For a discussion of attitudes, see the following: Harry C. Triandis, *Attitude and Attitude Change*, John Wiley and Sons (New York: 1971); *Attitudes Toward School, K-12*, Instructional Objectives Exchange, P.O. Box 24095, Los Angeles; *Attitudes Related to Tolerance*, Instructional Objectives Exchange, P.O. Box 24095, Los Angeles; and *How to Measure Attitudes* by Marlene E. Henerson, Lynn Lyons Morris, and Carol Taylor Fitz-Gibbon, Sage Publications (Beverly Hills: 1978).

perform on specific cognitive objectives or skills. On the other hand, if a program has explicit affective goals, and it is decided to determine how well those goals have been reached, it will be necessary to develop or purchase valid and reliable measures of student attitudes, feelings, values, or interests.

As an example, suppose that one of the purposes for testing is to determine how well students have mastered the cognitive objectives of a given course. If teachers have not previously identified objectives, or if the existing objectives are judged inadequate for some reason (perhaps they are stated in extremely general terms), it will be necessary to develop a list of instructional objectives.

In carrying out this task, two steps are involved:

First, it will be necessary to state all objectives in specific terms. One way to determine whether or not an objective is clearly and explicitly stated is to ask a number of persons to describe the types of test questions which measure the objectives. If there is a reasonable level of agreement among these individuals, then it is probably safe to assume the objective is well written.

Second, it may be necessary to limit the number of objectives to be assessed, simply because it may be unrealistic to test all possible learner outcomes. One approach for limiting the number is to have teachers rank them in importance. This way they will test only the objectives judged to be essential.

Another way to choose (limit) objectives for purposes of testing is on the basis of their position in a curricular hierarchy. For example, multiplication of whole numbers requires mastery of a number of prerequisite skills and is itself prerequisite to other higher level skills. For these two reasons, it would be a reasonable objective to test.

These two steps usually are difficult and time consuming because it is often difficult to get teachers to agree upon the number of instructional objectives to test, the exact wording of the objectives, the level of cognitive understanding, and what they consider to be satisfactory and/or exemplary performance.

3. Selecting or Developing a Test

After identifying both the purposes for testing and the information requirements of the testing program, selection or construction of appropriate test instruments is the next step. There are five types of tests from which to choose:

a. Norm-Referenced Tests
b. Objective-Referenced Tests
c. Customized Objective-Referenced Tests
d. Item Bank Tests
e. District-Developed Tests

The paragraphs which follow discuss the most important features of each of these options.

a. Norm-Referenced Tests[4]

If the chosen instrument is norm-referenced, it is essential that the basic characteristics of these tests be understood. A norm-referenced test is designed to sample a broad range of a subject matter area, but such tests do not typically present information about performance on specific instructional objectives. Norm-referenced tests do "sort" or spread students along a continuum or scale. The score a student receives on a norm-referenced test has meaning as it relates to the performances of other students. Thus, a percentile score of 65 means that a student has scored higher than 65 percent of the students who took the test when it was originally normed.

Typically, norm groups are state or national samples. Information about how a student (or group of students) compares to a norm group can provide a general idea of performance. However, such tests are not suitable for identifying specific strengths or weaknesses of students or for making judgments about the strengths or weaknesses of a program.

Popham makes the following observation about two basic weaknesses of a norm-referenced test:[5]

---

[4]Information about published tests can be found in Oscar Buros, *The Eighth Mental Measurements Yearbook*, Gryphon Press (Highland Park, New Jersey: 1978). In addition, the reader is advised to consult publications of the Center for the Study of Evaluation (CSE) at UCLA. For example, see the following: *CSE-RBS Test Evaluations: Tests of Higher-Order Cognitive, Affective and Interpersonal Skills* (Los Angeles: 1972); and *CSE Elementary School Test Evaluations* (Los Angeles: 1976).

[5]See W. James Popham, "An Approaching Peril: Cloud-Referenced Tests," *Phi Delta Kappan*, May 1974, Vol. 55, No. 9, pp 614-615.

> The first prominent defect of norm-referenced tests stems from their primary function, that is, to permit comparisons of an individual's performance with that of a norm group. . . .In order to permit an individual's performance to be referenced to the performance of a norm group, there must be a reasonable degree of response variance in scores on the test. In other words, to make fine-grained comparisons (e.g., Mary is at the 76th percentile; John is at the 74th), the test must produce a reasonable spread of scores. Test items which do not contribute substantially to the production of score variance are either modified or eliminated. . . . An ideal test item for maximizing response variance is one which only 50% of the test takers answer correctly. An item answered correctly by 80% or more of the students is typically jettisoned, since it doesn't do its share in spreading learners out. This means that in many cases the very items which deal with the concepts or skills considered most important to a field will, over time, be systematically eliminated from a norm-referenced achievement test. After several revisions, in fact, a standardized achievement test looks and functions very much like a standardized intelligence test. Such tests are relatively insensitive to instruction.
>
> A second basic problem with norm-referenced tests is that they are developed by test production firms which must service an entire nation, a nation with very diverse curricular preferences. If the test agencies spelled out, in unambiguous detail, exactly what it is their tests measure, many educators would find that the test's emphases did not match the local curriculum. They wouldn't buy the test. But test publishers subsist on the sales of their tests. They get out of this bind by describing their tests in very general terms, thereby picking up the classic Rorschach dividend of allowing people to see what they want to in an ink blot. This practice results in many educators using tests which are, in reality, not consonant with local curricular emphases. Mismatched measurement and instruction yields misleading data and spurious conclusions.

In summary, norm-referenced tests are useful to screen students for purposes of placement or guidance. They also allow one to make comparisons with an outside norm group. However, norm-referenced tests generally are inappropriate for purposes of diagnosing student strengths or weaknesses or for judging whether or not a program is effectively achieving its instructional objectives. In order to use a norm-referenced test for these purposes, it would be necessary to study in detail all test objectives and re-

lated items, and then to conclude that the test is, in fact, an appropriate instrument for a given set of objectives.[6]

b. Objective-Referenced Tests

An objective-referenced test is designed to gather information about student performance as it relates to specific instructional objectives.[7] The publisher should demonstrate that the items included on the test validly measure the stated objectives. An objective-referenced test may also provide norms (for example, state averages). However, this feature is secondary to its primary purpose: to provide information about student performance on a battery of objectives.

Choosing an objective-referenced test poses the question: Do the objectives tested match those emphasized in the local program? In the absence of a perfect match between local objectives and those listed on the test, the test may still be of use if only scores on selected objectives are considered. For example, if two of the test's ten objectives are not stressed in the curriculum, the student scores on these objectives can be disregarded in analysis and reporting.

c. Customized Objective-Referenced Tests

If district needs cannot be met by published tests and it is decided to develop tests, one option is to contract for customized objective-referenced tests. The usual procedure is to develop a list of objectives and present them to a testing company. In turn, the test publisher will provide appropriate test questions for each objective. If it is decided to use this approach, consult the contractor ahead of time because specific procedures differ from company to company.

There are at least two disadvantages associated with the use of customized objective-referenced tests. First, this type of service can be quite ex-

---

[6]Typically, norm-referenced tests will provide a list of objectives and questions keyed to those objectives. However, keep in mind that the primary intent of a norm-referenced test is to "sort" students, not to measure student performances on specific skills or objectives. Such global tests do not contain a sufficient number of items per objective to enable reliable decisions to be made about educational achievement.

[7]There is often confusion regarding the difference between an objective-referenced and criterion-referenced test. The basic difference is that the criterion-referenced test has pre-established levels of "acceptable" or "mastery" learning. An objective-referenced test does not have performance standards.

pensive. Second, because the test is designed to meet specific needs, it will not be possible to compare student scores with another group. If comparisons are an important component of a testing program, it will be necessary to use a test which has been widely administered and has norms associated with it.

d. Item Bank Tests

A second option for developing tests is to use an item bank. Ideally, an item bank refers to a computerized collection of items which are readily accessible to potential users. Such a system allows a user to select items from a variety of choices keyed to established objectives. Because the process of test construction is computerized, a functioning item bank allows teachers to develop tests with relative ease.

In addition to completely developed item banks, there are numerous collections of items which may be of use for test development. To use such collections, users must first match their objectives to those of the particular collection and then access the items keyed to the objectives of interest. The process in this instance is not computerized. As a result, the process of developing a test will probably be more tedious and time consuming.

e. District-Developed Tests

A third option is to have district personnel develop the tests. If this approach is selected, it is recommended that a basic textbook in tests and measurement be consulted.[9] Usually there is a section which deals with the principles of test construction. It is not practical to discuss the details of developing a test in a paper of this length.

Locally developed tests have great appeal because they can be designed to measure a district's specific instructional objectives. However, a major consideration is the significant amount of time and effort required to design a valid and reliable test instrument. For example, each test produced by the Wisconsin Pupil Assessment Program requires approximately two years to develop and involves between 50 and 100 persons. Time and effort is needed to develop objectives, write and revise test items, field test items, and pilot test items on a statewide sample.

Regardless of which test option is chosen, there are two related questions which require answers:

1. Who is to make the choice of a test or test items?

---

[9]For example, see Robert Ebel, *Essentials of Educational Measurement*, Prentice-Hall, (Englewood Cliffs, New Jersey: 1979).

2. What criteria are to be used in making the choice?

As mentioned previously, decision makers should be involved in choosing a test or test items. Without question, the judgment of teachers is needed because they have the subject-matter expertise which is crucial in reaching a final decision. However, there is another equally important reason for their involvement. If the testing program is to be effective, it must have the cooperation of teachers whose students are to be tested.

The answer to the second question, "What criteria are to be used in making the choice?," is more complex. The most important criterion must be whether or not the test will produce the needed information.

There are several additional criteria which might be used to judge potential tests, whether purchased or developed. These criteria are particularly applicable if it is decided to purchase a test from an outside source.

The following are among the other most important criteria:

1. *Validity:* A test which measures the established instructional objectives is valid. If a test is purchased, note how the test maker addresses the question of validity.

2. *Reliability:* A test is reliable if the scores of students are consistent. That is, if students receiving no training were to receive very different scores on the same test at different points in time, the test would have low reliability. All test publishers should address the question of test reliability in materials accompanying their tests. Typically, standard errors and reliability coefficients such as the Kuder-Richardson formula 21 or the Cronbach Alpha are included in the scoring services.

3. *Ease of Administration:* Are the directions provided for both the students and those who administer the test clear? If not, the results may be invalid and unreliable.

4. *Administration Time:* Is the time required to test students reasonable, or is an inordinately large or small amount of time required?

5. *Scoring Service:* What types of scoring services are available? What is the turn-around time? Does the scoring service provide the kinds of item data and test statistics needed?

6. *Costs:* What are the costs of test booklets, technical manuals, answer sheets, as well as the basic and optional scoring services?

7. *Special Features:* Does the test have any features which are particularly attractive? What are those features? How might they be used? Do they coincide with the original purpose for testing? What do they cost?

4. Administering the Test(s)

If a test is purchased, carefully read the administration instructions provided by the test publisher. In addition to these instructions, there are some general rules which should be followed in any testing situation. Among these are the following:

1. Be certain that all students understand the purpose for testing as well as the mechanics of the specific testing situation.
2. Organize all materials prior to the testing period in order to use time effectively.
3. Provide a proper testing environment: adequate lighting and ventilation, minimal distractions, etc.
4. Conduct multi-part tests or lengthy tests in separate testing sessions.
5. Avoid test administration on days before vacation or at times prior to special school events (e.g., the hour before an all-school assembly would probably be a poor time to test).
6. Make some provision for students who are absent to take the test.

5. Scoring and Analyzing Test Results

It is not possible to adequately discuss scoring or analysis of test results in a brief paper. Scoring and analysis, however, represent essential components of a testing program. These aspects of a testing program involve more than making arrangements for scoring and then tabulating the test scores into summary form. The data need to be organized and studied in a systematic manner. It is important that a person with training and experience in the analysis of test results be involved in this phase of a testing program.

It also is advisable that additional kinds of information be used in the analysis of test results. For example, student scores on other tests or consultations with teachers, as well as with the person who administered the test, may provide information which is of value in understanding the results of a specific test. Above all, it should be remembered that no single test mea-

sures all that students learn or experience. In reaching any conclusions or making any recommendations, this point must be kept in mind.

6. Preparing and Publishing a Report

A report should address five essential topics[9]

1. What are the purposes for testing?
2. Who was tested, what areas were tested, and when did testing occur?
3. Why was a particular published test chosen, or, if tests were locally developed, what was the process of developing the test?
4. What do the results show?
5. What are the conclusions and recommendations?

While all reports should deal with each of the above, the specific emphasis and amount of detail will vary, depending on the target audience. Most likely, it will be necessary to prepare different kinds of reports, simply because the information desired by one group is unlikely to be the same as that desired by another.

In preparing an oral or written report, the following are recommended:

1. Keep your audience in mind at all times. Limit the presentation by focusing it towards the kind of information they can use. Emphasize the important points. Don't present too much detail.
2. Make your report interesting.
3. If the report is oral, allow time for questions. If the report is written, encourage feedback.
4. Use graphs, charts, tables, and other visual aids.
5. Exercise caution in making comparisons among schools or classes on test scores. If comparisons are to be included in the report, recognize that there are considerable risks involved.
6. Emphasize the fact that a test measures only part of what takes place on a daily basis. Under no conditions should the scores

[9]An excellent summary of reporting test results can be found in the Program Evaluation Kit by Lynn Lyons Morris and Carol Taylor Fitz-Gibbon, op. cit. See the publication entitled *How to Present an Evaluation Report.*

from a single test (or even scores from a series of tests) be used to judge the total educational program.

Analysis of test results should provide some ideas for improvement; these ideas may be included in the report. Some of the ideas will have been considered during planning, when decisions to be based on findings were discussed. Others may occur to the report writer after examination of the data. Recommendations can be included within the final report, or if preferred, they can be presented separately. Recommendations might include initiating particular teaching strategies, holding inservice programs, acquiring new materials and textbooks, giving additional emphasis to selected topics, modifying curriculum and objectives, and providing additional help to certain students. Decisions of this type usually require a background in such areas as learning theory, curriculum, and testing. Persons with expertise in any or all of these areas represent ideal resources for this stage of a testing program.

7. Providing for Follow-up

The basic reason for initiating a testing program of any type is to gather information to make educational decisions. If the test data provide the necessary information, it should be possible to make decisions which are effective. What is needed as a final step in a testing program, then, is follow-up, where it is determined whether or not the decisions or recommendations have had the desired effect. Additional testing and discussions with teachers and students represent two ways to determine whether or not the changes have been of value. In effect, the follow-up is the first step in planning for the next cycle of testing.

SOME RECOMMENDATIONS FOR TESTING PROGRAMS

There is no single model or approach to establishing a viable testing program. However, listed below are six recommendations which should be helpful in establishing a testing program which is both effective and positively received.

1. Testing should be conducted in a positive manner. A testing program should be intended to improve the quality of both educational programs and instruction. If weaknesses are indicated, emphasis should be on correction, not censure.

2. All testing efforts should be realistic in scope. It is not possible to measure all that a student experiences in school. What this means is

that there is a need to be selective throughout the entire process of testing.

3. It should be the goal of each subject area to build a testing component into its instructional program. For example, a final examination in a course can be designed not only to provide a basis for student grades, but also to provide teachers with helpful information to improve instruction and to give district-wide programmatic information.

4. It is important to have consensus in all testing efforts. Teachers, administrators, board members, and parents should all participate in identifying course objectives, developing or selecting tests, analyzing data, and recommending changes.

5. If existing tests are adopted, the particular features of those tests should be well understood. Above all, the tests selected should reflect the district's instructional objectives. If they do not, the test scores will certainly be invalid and may be unreliable.

6. Test information collected regularly over a period of time is desirable. For this reason, it is recommended that a testing program be carried out over a period of years. A one-time testing effort is of relatively little value.

PART II

TESTING IN WISCONSIN PUBLIC SCHOOL DISTRICTS

In January 1980, a questionnaire was sent to Wisconsin's 433 public school districts by the Wisconsin Pupil Assessment Program of the Department of Public Instruction. The questionnaire, completed by 303 (70%) of the districts, was designed to answer two questions about district testing programs:

1. What tests are used in Wisconsin public schools?
2. What use is made of test results?

Summary of Tests Used in Wisconsin Public School Districts

In response to the questionnaire, Wisconsin school districts indicate that tests currently used fall into three categories:[10]

1. Tests purchased from commercial testing companies
2. Tests purchased from the Wisconsin Pupil Assessment Program
3. District-developed tests

1. Commercial Tests

In terms of commercial tests, the questionnaire asked each district what tests were currently being used, or had been used during the previous three years.[11]

In addition to naming the tests used, districts were asked to indicate the grade levels of test use. The choices were K-5, 6-8, and 9-12. The results of this section of the questionnaire are shown in Table 1.

Besides showing the test name, test publisher, and grade usage, Table 1 indicates the type of test, either norm- or objective-referenced.[12] It is important not to equate the frequency of test use by districts, as shown in this table, with the number of students actually tested. For example, Milwaukee uses several commercial tests in its citywide testing program. One of these tests, the Differential Aptitude Tests, is shown in Table 1 to have a relatively modest frequency of use throughout the state. However, because it is used in the state's largest district, it may, in fact, be one of the tests taken by the greatest number of students in Wisconsin.

---

[10] It should be noted that several Wisconsin school districts use item collections offered by CESA agencies. However, use of this service was not indicated in the questionnaire under a separate category.

[11] Respondents also were asked to list tests used in their districts, but not appearing on the list. More than 150 additional commercial tests were listed. However, only six of these were reported ten or more times: Armed Services Vocational Aptitude Battery (16 times), Cognitive Abilities Test (14 times), KeyMath Diagnostic Arithmetic Test (14 times), Peabody Individual Achievement Test (13 times), Wechsler Intelligence Scale for Children (10 times) and Woodcock Reading Mastery Tests (16 times).

[12] The distinction between norm-referenced and objective-referenced tests is discussed in Part I of this paper.

TABLE 1
COMMERCIAL TESTS USED IN WISCONSIN PUBLIC SCHOOL DISTRICTS

Name of Test	Publisher	Type of Test	Number of Districts Reporting Use		
			K-5	6-8	9-12
Achievement Series: SRA Assessment Series	Science Research Associates	NRT	49	48	34
American School Achievement Tests	Bobbs Merrill Co.	NRT	0	0	0
Analysis of Skills	Scholastic Testing Service	NRT	1	0	0
California Achievement Tests (CAT)	CTB/McGraw-Hill	NRT	32	26	18
Comprehensive Tests of Basic Skills (CTBS)	CTB/McGraw-Hill	NRT	18	23	24
Cooperative School and College Ability Tests	Addison-Wesley	NRT	1	2	17
Differential Aptitude Tests	Psychological Corporation	NRT	1	10	20
ERB Comprehensive Testing Program	Educational Records Bureau	NRT	0	0	0
Gates-MacGinitie Reading Tests	Riverside Publishing	NRT	106	88	47
Henmon-Nelson Tests of Mental Ability	Riverside Publishing	NRT	43	57	95
Iowa Tests of Basic Skills (ITBS)	Riverside Publishing	NRT	89	88	24
Iowa Tests of Educational Development (ITED)	Science Research Associates	NRT	3	5	32
Lorge-Thorndike Intelligence Tests	Riverside Publishing	NRT	27	25	8
Metropolitan Achievement Tests	Psychological Corporation	NRT	79	39	20
Metropolitan Readiness Tests	Psychological Corporation	NRT	112	3	1
Objectives-Referenced Bank of Items and Tests	CTB/McGraw-Hill	ORT	1	1	0
Otis-Lennon Mental Ability Test	Psychological Corporation	NRT	81	61	33
Prescriptive Reading Inventory (PRI)	CTB/McGraw-Hill	NRT	8	3	1
Sequential Tests of Educational Progress (STEP)	Addison-Wesley	NRT	15	33	78
Stanford Achievement Tests	Psychological Corporation	NRT	84	73	23
Stanford Test of Academic Skills	Psychological Corporation	NRT	6	7	19
STS Educational Development Series	Scholastic Testing Service	NRT	6	7	10
Tests of Academic Progress	Riverside Publishing	NRT	5	2	15
School Curriculum Objective-Referenced Evaluation (SCORE)	Westinghouse Research Corporation	ORT	0	4	1
SRA Criterion-Referenced Measured Program	Science Research Associates	ORT	6	6	7
Stanford Diagnostic Mathematics Test	Psychological Corporation	NRT	21	19	8
Stanford Diagnostic Reading Test	Psychological Corporation	NRT	44	38	16
Wide Range Achievement Test	Guidance Associates	NRT	68	61	52

Two conclusions can be drawn from the data shown in Table 1. First, in nearly all cases, the commercial tests used by Wisconsin school districts are norm-referenced. Second, the commercial tests are produced by relatively few test publishers, as shown in Table 2. In this case, the five most frequently used commercial tests for each of the three grade level categories is shown.

TABLE 2
FIVE MOST FREQUENTLY USED COMMERCIAL TESTS IN GRADE LEVELS K-5, 6-8, AND 9-12

Test	Publisher	Number of Districts Reporting Use
	Grades K-5	
Metropolitan Readiness Tests	Psychological Corporation	112
Gates-MacGinitie Reading Tests	Riverside Publishing	105
Iowa Tests of Basic Skills	Riverside Publishing	89
Stanford Achievement Tests	Psychological Corporation	84
Otis-Lennon Mental Ability Test	Psychological Corporation	81
	Grades 6-8	
Iowa Tests of Basic Skills	Riverside Publishing	88
Gates-MacGinitie Reading Tests	Riverside Publishing	88
Stanford Achievement Tests	Psychological Corporation	73
Otis-Lennon Mental Ability Test	Psychological Corporation	61
Wide Range Achievement Test	Guidance Associates	61
	Grades 9-12	
Henmon-Nelson Tests of Mental Ability	Riverside Publishing	95
Sequential Tests of Educational Progress	Addison-Wesley Publishers	78
Wide Range Achievement Test	Guidance Associates	52
Gates-MacGinitie Reading Tests	Riverside Publishing	47
Achievement Series: SRA Assessment Survey	Science Research Associates	34

2. Tests Purchased from the Wisconsin Pupil Assessment Program

Based on responses to the survey, tests developed by the Wisconsin Pupil Assessment Program have been used by approximately 20 percent of Wisconsin public school districts.[13] The mathematics tests are the most

[13]Note that these figures represent district use at some time during the period, 1978-1980. For any single year, the percentage would be lower.

frequently used, followed by the reading tests. Relatively few districts have used tests in geometry or economic understanding.

All tests developed by the Wisconsin Pupil Assessment Program are objective-referenced. Each has gone through a systematic developmental process. Committees consisting of teachers, parents, employers, and others from throughout the state have assumed a central role in identifying the objectives or skills that students should be expected to know at the various grade levels. Similar committees have played a major role in writing or selecting test questions to measure established objectives.

Although tests developed by the Wisconsin Pupil Assessment Program are designed for purposes of statewide assessment, each is made available to public and private schools through a local option testing service.[*] As part of this service, test users are provided with administration instructions, test scoring arrangements, and reporting services. Table 3 indicates the number of Wisconsin public school districts using the local option, from 1978 through 1980.

TABLE 3
NUMBER OF PUBLIC SCHOOL DISTRICTS
USING TESTS DEVELOPED BY
THE WISCONSIN PUPIL ASSESSMENT PROGRAM
1978-1980

		Grade Levels Tested*		
Subject Area	**Grades**	**K-5**	**6-8**	**9-12**
Mathematics	4, 8, 12	67 (22%)**	81 (27%)	69 (23%)
Geometry	4, 8, 12	14 (5%)	15 (5%)	19 (6%)
Reading	4, 8, 12	61 (20%)	61 (20%)	46 (15%)
Economic Understanding	12	-	-	39 (13%)

*Although tests are developed for statewide assessment at specific grade levels, they may be used at a variety of grade levels.

**These figures refer to the percent of K-5, 6-8, and 9-12 districts responding to the survey.

[*]For more information about the local option, contact the Wisconsin Pupil Assessment Program, 125 S. Webster Street, Madison, Wisconsin 53702. Request the *Testing Services Manual.*

3. District-Developed Tests

Most Wisconsin school districts surveyed indicate dependence on commercially produced norm-referenced tests or Wisconsin Pupil Assessment Program objective-referenced tests. Relatively few districts (perhaps 10-15 percent) have developed their own test instruments for purposes of district-wide testing. The numbers and subject areas are shown in Table 4.

TABLE 4
NUMBER OF PUBLIC SCHOOL DISTRICTS USING TESTS DEVELOPED BY THE DISTRICT
1978-1980

Subject Area	Grade Levels Tested K-5	6-8	9-12
Mathematics	39 (13%)*	51 (17%)	41 (14%)
Reading	30 (10%)	26 (9%)	17 (6%)
Writing	8 (3%)	12 (4%)	18 (6%)
Others	19 (6%)	19 (6%)	6 (2%)

*These figures refer to the percent of K-5, 6-8, and 9-12 districts responding to the survey.

Generally, districts have developed tests in three areas: mathematics, reading, and writing. About 6 percent of districts responding listed tests in subjects other than these three. There are perhaps two basic reasons for the focus on these subjects. First, reading, mathematics, and writing represent fundamental or basic skill areas. Thus, the emphasis given to these subjects is to be expected. Second, developing valid and reliable objective-referenced tests in mathematics, reading, and writing is generally an easier task than developing a test in such subjects as science or social studies.

District Use of Test Results

The survey results suggest that tests, regardless of source, tend to be used for essentially the same purposes. Generally, the most frequently stated use of test results is to diagnose student strengths and weaknesses.

Other frequent uses are to identify curricular strengths and weaknesses, to assist in guidance and counseling, and to place pupils. In contrast, evaluation of teachers is seldom reported as a use of test results.

Tests developed by the Wisconsin Pupil Assessment Program or private testing companies are also used in many instances to compare students to an outside group. District-developed tests, of course, cannot be used for this purpose. The specific breakdown of test usage is shown in Table 5.

Table 5

REPORTED USES OF TESTS IN WISCONSIN PUBLIC SCHOOL DISTRICTS

	Number of Districts Using Various Tests		
Use of Test Results*	**Commercial Tests**	**Wis. Pupil Assessment**	**District-Developed**
Identification of curricular strengths and weaknesses	225	74	65
Promotion or graduation	15	2	22
Evaluation of teachers	7	1	1
Pupil placement	205	19	60
Diagnosis of students' strengths or weaknesses	253	54	74
General idea of how students compare with others	270	72	NA
Research	98	16	16
Guidance and counseling	248	45	36

*Note that a given test may have multiple uses.

CONCLUSIONS AND RECOMMENDATIONS

School districts increasingly are being held accountable to demonstrate the effectiveness of the educational process. One direct consequence of the current demands for accountability is that Wisconsin public school districts are doing a great deal of testing. However, it is apparent from the survey results that there are numerous inconsistencies between the uses of tests and the purposes for which the tests were designed.

Responses to the survey indicate that the majority of Wisconsin public school districts purchase tests, rather than develop their own district-wide test instruments. Furthermore, the survey indicates that the purchased tests are overwhelmingly norm-referenced. Norm-referenced tests are most appropriate for purposes of comparing students to other students,

and to some extent, for pupil placement, research, and guidance and counseling. The data show that many districts do use norm-referenced tests for these reasons.

At the same time, there are many instances of districts using norm-referenced tests for identification of curricular strengths and weaknesses, promotion or graduation, diagnosis of individual student strengths and weaknesses, and in a few cases, for evaluation of teachers. These uses of norm-referenced tests are not recommended, essentially because norm-referenced tests were not designed for such purposes. Objective-referenced tests are more appropriate choices.

The available data do not explain why Wisconsin public school districts have chosen to use commercially produced norm-referenced tests as they do. However, some possible explanations include the following: (1) the tests have been used for a number of years and, therefore, are familiar to decision-makers; (2) the tests are convenient to use because they require no time and effort for development, piloting, and revision; (3) the tests are available at a known cost; (4) the tests cover a wide range of content and grade levels; (5) the tests usually provide norms for a general comparison of local achievement; (6) the tests have a high reliability and small measurement error; and (7) there may be a misunderstanding of the essential differences between norm-referenced and objective-referenced tests.

In light of the survey results, it is apparent that school districts need to review what they are doing in the area of testing. This means a careful examination of the purposes for testing. It also means giving attention to all components of the existing testing program, with a special look at whether or not tests presently being used provide the necessary information to make important educational decisions. If there are weaknesses in the testing program, these weaknesses should be corrected. In some instances, this may require a revamping of the existing program.

School districts have both the capacity and the responsibility to assume a leadership role in the areas of testing and accountability. This responsibility includes establishing a testing program that is carefully planned and implemented. It includes informing the community about the purposes for testing and design of the district's testing program. It also includes informing the community about the benefits and limitations of any testing program. A strong school system requires community support. The willingness of school districts to assume a leadership role in testing and accountability will enhance that support.